Docket 82600JDP
Customer No. 01333

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

In re Application of	Group Art Unit: 2179
Alexander C. Loui, et al.	Examiner: Sara M. Hanne
VIDEO STRUCTURING BY PROBABILISTIC MERGING OF VIDEO SEGMENTS	
Serial No. 09/927,041	
Filed 09 August 2001	

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

APPEAL BRIEF PURSUANT TO 37 C.F.R. 41.37 and 35 U.S.C. 134

I. Table Of Contents

I. Table Of Contents i
II. Real Party In Interest 1
III. Related Appeals And Interferences 2
IV. Status Of The Claims 3
V. Status Of Amendments 4
VI. Summary of Claimed Subject Matter 5
VII. Grounds of Rejection to be Reviewed on Appeal 14
VIII. Arguments 15
IX. Conclusion 38
X. Appendix I - Claims on Appeal 39
XI. Appendix II - Evidence 52
XII. Appendix III – Related Proceedings 53

APPELLANT'S BRIEF ON APPEAL

Appellants hereby appeal to the Board of Patent Appeals and Interferences from the Examiner's Final Rejection of claims 1 and 3-29 which was contained in the Office Action mailed June 5, 2006..

A timely Notice of Appeal was filed November 3, 2006.

II. Real Party In Interest

As indicated above in the caption of the Brief, the Eastman Kodak Company is the real party in interest.

III. Related Appeals And Interferences

No appeals or interferences are known which will directly affect or be directly affected by or have bearing on the Board's decision in the pending appeal.

IV. Status Of The Claims

Claim 2 has been canceled. Claims 1, and 3 – 29 stand rejected. Claims 1 and 3 – 29 are the subject of this appeal.

Appendix I provides a clean, double spaced copy of the claims on appeal.

V. Status Of Amendments

No amendments to the claims were made after the final rejection. Remarks were made, however, by Appellants in a Response After Final dated October 3, 2006. An Advisory Action mailed October 31, 2006 indicated that such remarks did not place this application in condition for allowance.

VI. Summary of Claimed Subject Matter

Independent claim 1 reads as follows;

1. A method for structuring video by probabilistic merging of video segments, said method comprising the steps of:

a) obtaining a plurality of frames of unstructured video (Fig. 1, 8; Spec. page 7, lines 13 – 15);

b) generating video segments from the unstructured video by detecting shot boundaries based on color dissimilarity between consecutive frames (Fig. 1, 10; Spec. page 7, lines 19 – 25);

c) extracting a feature set by processing pairs of said segments, said extracting generating an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments, said inter-segment temporal relationship feature including metrics of temporal separation between the segments of the respective said pair and accumulated duration of the segments of the respective said pair (Fig. 1, 12; Spec page 7, line 26 – page 8, line 10); and

d) merging video segments with a merging criterion that applies a probabilistic analysis to the features of the feature set, thereby generating a merging sequence representing the video structure (Fig. 1, 14; Spec. page 8, lines 11 – 23).

Independent claim 7 reads as follows:

7. A method for structuring video by probabilistic merging of video segments, said method comprising the steps of:

a) obtaining a plurality of frames of unstructured video (Fig. 1, 8; Spec. page 7, lines 13 – 15);

b) generating video segments from the unstructured video by detecting shot boundaries based on color dissimilarity between consecutive frames (Fig. 1, 10; Spec. page 7, lines 19 – 25);

c) extracting a feature set by processing pairs of said segments, said extracting generating an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments (Fig. 1, 12; Spec page 7, line 26 – page 8, line 10); and

d) merging video segments with a merging criterion that applies a probabilistic analysis to the features of the feature set, thereby generating a merging sequence representing the video structure (Fig. 1, 14; Spec. page 8, lines 11 – 23);

wherein step d) comprises the steps of:

generating parametric mixture models to represent class-conditional densities of inter-segment features of the feature set, said parametric mixture models being statistical models; and

applying the merging criterion to the parametric mixture models.

Independent claim 10 reads as follows:

10. A computer storage medium having instructions stored therein for causing a computer to perform the acts of:

generating video segments from unstructured video by detecting shot boundaries based on color dissimilarity between consecutive frames (Fig. 1, 10; Spec. page 7, lines 19 – 25);

extracting a feature set by processing pairs of segments, said extracting generating an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments (Fig. 1, 12; Spec page 7, line 26 – page 8, line 10); and

merging video segments with a merging criterion that applies a probabilistic analysis to the features of the feature set, thereby generating a merging sequence representing the video structure (Fig. 1, 14; Spec. page 8, lines [illegible] 23);

wherein said merging further comprises the steps of:

generating statistical models of the feature set; and

applying the merging criterion to the statistical models.

Independent claim 11 reads as follows:

11. A method for structuring video by probabilistic merging of video segments, said method comprising the steps of:

a) obtaining a plurality of frames of unstructured video (Fig. 1, 8; Spec. page 7, lines 13 – 15);

b) generating video segments from the unstructured video by detecting shot boundaries based on color dissimilarity between consecutive video frames (Fig. 1, 10; Spec. page 7, lines 19 – 25);

c) extracting a feature set by processing pairs of segments, said extracting generating an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments (Fig. 1, 12; Spec page 7, line 26 – page 8, line 10);

d) generating a parametric mixture model of the inter-segment features comprising the feature set, said parametric mixture model being a statistical model (Fig. 1, 14; Spec. page 8, lines 11 – 23); and

e) merging video segments with a merging criterion that applies a probabilistic Bayesian analysis to the parametric mixture model, thereby generating a merging sequence representing the video structure (Fig. 1, 14; Spec. page 8, lines 11 – 23).

Independent claim 17 reads as follows:

17. A computer storage medium having instructions stored therein for causing a computer to perform acts for structuring video by probabilistic merging of video segments, the acts including:

obtaining a plurality of frames of unstructured video (Fig. 1, 8; Spec. page 7, lines 13 – 15);

generating video segments from the unstructured video by detecting shot boundaries based on color dissimilarity between consecutive video frames (Fig. 1, 10; Spec. page 7, lines 19 – 25);

extracting a feature set by processing pairs of segments, said extracting generating an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of cach said pair of segments (Fig. 1, 12; Spec page 7, line 26 – page 8, line 10);

generating a parametric mixture model of the inter-segment features comprising the feature set, said parametric mixture model being a statistical model (Fig. 1, 14; Spec. page 8, lines 11 – 23); and

merging video segments with a merging criterion that applies a probabilistic Bayesian analysis to the parametric mixture model, thereby generating a merging sequence representing the video structure (Fig. 1, 14; Spec. page 8, lines 11 – 23).

Independent claim 18 reads as follows:

18. A method for structuring video by probabilistic merging of video segments, said method comprising the steps of:

a) obtaining a plurality of frames of unstructured video (Fig. 1, 8; Spec. page 7, lines 13 – 15);

b) generating video segments from the unstructured video by detecting shot boundaries based on color dissimilarity between consecutive video frames (Fig. 1, 10; Spec. page 7, lines 19 – 25);

c) extracting a feature set by processing pairs of segments, said extracting generating an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments (Fig. 1, 12; Spec page 7, line 26 – page 8, line 10);

d) merging adjacent video segments with a merging criterion that applies a probabilistic Bayesian analysis to parametric mixture models derived from the feature set, said parametric mixture models being statistical models, thereby generating a merging sequence (Fig. 1, 14; Spec. page 8, lines 11 – 23); and

e) representing the merging sequence in a hierarchical tree structure (Fig. 1, 16; Spec. page 8, lines 24 – 28).

Independent claim 20 reads as follows:

20. A computer storage medium having instructions stored therein for causing a computer to perform probabilistic merging of video segments, said instructions performing the acts of:

a) obtaining a plurality of frames of unstructured video (Fig. 1, 8; Spec. page 7, lines 13 – 15);

b) generating video segments from the unstructured video by detecting shot boundaries based on color dissimilarity between consecutive video frames (Fig. 1, 10; Spec. page 7, lines 19 – 25);

c) extracting a feature set by processing pairs of segments, said extracting generating an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments ;

d) merging adjacent video segments with a merging criterion that applies a probabilistic Bayesian analysis to parametric mixture models derived from the feature set, said parametric mixture models being a statistical models, thereby generating a merging sequence (Fig. 1, 14; Spec. page 8, lines 11 – 23); and

e) representing the merging sequence in a hierarchical tree structure (Fig. 1, 16; Spec. page 8, lines 24 – 28).

Independent claim 21 reads as follows:

21. A method for structuring video by probabilistic merging of video segments, said method comprising:

generating video segments from an unstructured plurality of video frames by detecting shot boundaries based on color dissimilarity between consecutive frames (Fig. 1, 10; Spec. page 7, lines 19 – 25);

extracting a feature set by processing pairs of segments, said extracting generating an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments, said inter-segment temporal relationship feature including metrics of temporal separation between the segments of the respective said pair and accumulated duration of the segments of the respective said pair (Fig. 1, 12; Spec page 7, line 26 – page 8, line 10; page 11, line 30 – page 13, line 10);

merging the video segments with a merging criterion that applies a probabilistic analysis to the feature set, thereby generating a merging sequence representing the video structure, the merging being independent of any empirical parameter determination (Fig. 1, 14; Spec. page 8, lines 11 – 23); and

generating a hierarchy with the merged video segments, the hierarchy having a merging sequence represented by a binary partition tree (Fig. 1, 16; Spec. page 8, lines 24 – 28).

Independent claim 27 reads as follows:

27. A method for structuring video by probabilistic merging of video segments, said method comprising the steps of:

generating video segments from a plurality of frames of unstructured video by detecting shot boundaries based on color dissimilarity between consecutive frames(Fig. 1, 10; Spec. page 7, lines 19 – 25);

computing an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments, said inter-segment temporal relationship feature including metrics of temporal separation between the segments of the respective said pair and accumulated duration of the segments of the respective said pair (Fig. 1, 12; Spec page 7, line 26 – page 8, line 10; page 11, line 30 – page 13, line 10); and

d) merging video segments with a merging criterion that applies a probabilistic analysis to said features, thereby generating a merging sequence representing the video structure (Fig. 1, 14; Spec. page 8, lines 11 – 23).

Independent claim 29 reads as follows:

29. A method for structuring video by probabilistic merging of video segments, said method comprising the steps of:

obtaining a plurality of frames of unstructured video (Fig. 1, 8; Spec. page 7, lines 13 – 15);

generating video segments from the unstructured video by detecting shot boundaries based on color dissimilarity between consecutive frames (Fig. 1, 10; Spec. page 7, lines 19 – 25);

extracting an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments, said extracting of said inter-segment temporal relationship feature of each said pair of segments including determining a number of frames separating the respective said pair of segments and determining an accumulated number of frames in said segments of the respective said pair of segments(Fig. 1, 12; Spec page 7, line 26 – page 8, line 10; page 11, line 30 – page 13, line 10); and

merging video segments with a merging criterion that applies a probabilistic analysis to the features of the feature set, thereby generating a merging sequence representing the video structure (Fig. 1, 14; Spec. page 8, lines 11 – 23).

VII. Grounds of Rejection to be Reviewed on Appeal

The following issues are presented for review by the Board of Patent Appeals and Interferences:

(1) The rejection of claims 1, 3 – 8, 10, and 23 – 29 under 35 USC 103(a) as being unpatentable over Qian et al (US Patent 6,721,454) in view of Ratakonda (US Patent 5,956,026).

(2) The rejection of claims 9, and 11 – 22 under 35 USC 103(a) as being unpatentable over Qian et al. (US Patent 6,721,454), in view of Ratakonda (US Patent 5,956,026), and further in view of Qian et al. (US Patent 6,616,529).

VIII. Arguments

A. Claims 1, 3 – 8, 10, and 23 – 29 are patentable under 35 USC 103(a) over Qian et al. US Patent 6721454 (Qian 454), Ratakonda US Patent 5956026, further in view of Qian et al. US Patent 6616529 (Qian 529).

Regarding Independent Claims 1 and 7, the rejection inconsistently interprets "shots" described in the Qian 454 Patent.

Claim 1 requires, among other things, "generating video segments from the unstructured video by detecting shot boundaries". As allegedly teaching this features, the rejection refers to Col. 3, lines 42-43 of Qian 454. See the top of Page 3 of the Office Action. This and adjacent portions of the Qian 454 Patent state:

> A video sequence includes one or more scenes which, in turn, include one or more video shots. A shot comprises a plurality of individual frames of relatively homogeneous content. At the first level of the [Qian 454's] technique 4, the boundaries of the constituent shots of the sequence are detected 6. A color histogram technique may be used to detect the boundaries of the shots. The difference between the histograms of two frames indicates a difference in the content of those frames. When the difference between the histograms for successive frames exceeds a predefined threshold, the content of the two frames is assumed to be sufficiently different that the frames are from different video shots. Other known techniques could be used to detect the shot boundaries.

Col. 3, lines 36-50.

Accordingly, Appellants understand this portion of the rejection to take the position that the shots defined by Qian 454's above-cited boundary detection process correspond to the video segments recited in Claim 1.

However, Claim 1 further requires "extracting a feature set by processing pairs of said segments". (underline added for emphasis). In order to teach this limitation, and in view of the above interpretation of the Qian 454 Patent, such Patent would have to teach extracting a feature set by processing

pairs of the shots defined by the boundary detection process described at col. 3, lines 36-50.

However, because the Qian 454 Patent does not include such a teaching or suggestion, the rejection changes its definition of "shot" according to the Qian 454 Patent in an attempt to meet the limitations of Claim 1. Instead of referring to the shots defined by the boundary detection process described at col. 3, lines 36-50, the rejection refers to a different aspect of the Qian 454 Patent pertaining to adding additional shot boundaries based upon calculated motion changes between frames. See the top of Page 3 of the Office Action, referring to Col. 3, lines 59-61 of the Qian 454 Patent. The relevant portions of the Qian 454 Patent state:

> In addition to the shot boundaries detected in the video sequence, shot boundaries may be forced or inserted into the sequence whenever the global motion of the content changes. As a result, the global motion is relatively homogeneous between the boundaries of a shot . . . . At the first level 4 of the technique, the global motion of the video content is estimated 8 for each pair of frames in a shot.

Col. 3, lines 51-61.

Apparently, the rejection interprets this motion change evaluation process to involve a two step process: (1) calculating motion between each pair of frames, and (2) when there is a 'jump' in two consecutive calculated motions, an additional shot boundary is inserted. In contrast to Claim 1, step (1) does not involve "segments", because a segment includes more than one frame. Accordingly, the rejection appears to be referring to step (2) to allegedly teach extracting a feature set by processing pairs of said segments, as required by Claim 1. In this regard, it appears that the rejection is reasoning that step (2) involves processing pairs of segments because it involves comparing two calculated motions, each calculated motion having been derived from a pair of frames. It follows then, according to the apparent reasoning of the rejection, that a "segment" is a pair of frames according to Col. 3, lines 59-61 of Qian 454.

If Appellants' above-described understanding of the rejection is correct, it improperly changes the definition of "segments" from (a) shots defined by the boundary detection process described at col. 3, lines 36-50 to (b) each and every pair of frames. Appellants respectfully submit that such a change in definition is improper because claim 1 requires "extracting a feature set by processing pairs of said segments" where --said segments-- refers to the segments "generat[ed] . . . by detecting shot boundaries". Accordingly, the definition of "segments" given to Claim 1's "generating" step must be consistent with the definition of "segments given to Claim 1's "extracting" step. Therefore, Appellants respectfully submit, based upon the above-described interpretation of the Qian 454 Patent, such Patent would have to teach extracting a feature set by processing pairs of the shots defined by the boundary detection process described at col. 3, lines 36-50. However, the Qian 454 Patent is not understood to and has not been cited to provide such a teaching or suggestion.

The other references of record are not cited as teaching or suggesting these features of Claim 1. Accordingly, Claim 1 is respectfully submitted to be patentable. Independent Claim 7 includes the same or similar features as those discussed above in connection with Claim 1, and is submitted to be patentable for at least the same reasons. The claims that depend from Claims 1 or 7 also are submitted to be patentable for at least the same reasons.

Regarding All Claims: The Qian 454 Patent does not Teach or Suggest Merging Video Segments with a Merging Criteria that Applies a Probabilistic Analysis to the Features of the Feature Set.

Claim 1 requires "extracting a feature set by processing pairs of said segments" and "merging video segments with a merging criterion that applies a probabilistic analysis to the features of the feature set." The rejection is unclear as to what exactly it is referring to in the Qian 454 Patent as allegedly disclosing a "feature set" according to Claim 1. In particular, the rejection first cites col. 3, lines 59-61 of the Qain 454 Patent to allegedly teach "extracting a [feature] set by processing pairs of segments". See the top of page 3 of the Office Action. This portion of the Qain 454 Patent refers to the motion calculations described earlier:

> At the first level 4 of the [Qain 454] technique, the global motion of the video content is estimated 8 for each pair of frames in a shot.

Col. 3, lines 59-61.

Accordingly, it appears that the rejection is implying that the motion estimation for each pair of frames is a "feature set" according to Claim 1. However, the motion estimation is calculated for each pair of frames, and as discussed earlier, a frame is not a segment. Accordingly, Appellants respectfully submit that the motion estimation calculation for each pair of frames cannot teach or suggest "extracting a feature set by processing pairs of said segments," as required by Claim 1.

To reason further, however, the rejection may be implying that the Qian 454 Patent's comparison of different motion estimations to determine where a 'jump' in motion estimations occurs is an 'extracted feature set' according to Claim 1. If this were the case, however, the Qian 454 Patent would also have to teach or suggest that such feature set (i.e., results of comparisons of motion estimations) would be used to "merge video segments with a merging criterion that applies a probabilistic analysis to the" results of the comparisons of the motion estimations, in order to teach the merging step of Claim 1. However, the Qian 454 Patent includes no such teaching and, in fact, teaches the opposite. As described at col. 3, lines 51-55, the results of the comparisons of the motion estimations are used to further <u>split</u> shots, not merge shots.

> In addition to the shot boundaries detected in the video sequence, shot boundaries may be forced or inserted into the sequence whenever the global motion of the content changes.

Col. 3, lines 51-55.

Accordingly, Appellants respectfully submit that the Qian 454 Patent's teaching of comparing motion estimations in order to determine where to insert additional shot boundaries does not teach or suggest "merging video segments with a merging criterion that applies a probabilistic analysis to the features of the feature set," as required by Claim 1.

In view of the above, Appellants respectfully submit that the Qian 454 Patent at least does not teach or suggest "extracting a feature set by processing pairs of said segments" and "merging video segments with a merging criterion that applies a probabilistic analysis to the features of the feature set", as required by Claim 1.

In regard to the Ratakonda reference, the rejection indicates:

> In addition, Ratakonda further teaches the generation of inter-segment color dissimilarity feature and inter-segment temporal relationship feature of each pair of segments (Figures 1, 5 and corresponding text).

See the bottom of Page 3 of the Office Action.

Figure 1 of Ratakonda does not show pairs of segments and is not discussed in terms of "each pair of segments". Figure 5 shows a finest level 76, which includes five keyframes. The rejection is apparently based upon an assumption that each of the keyframes in Figure 5 of Ratakonda corresponds to a shot (segment). This assumption has appeared to have some support. Ratakonda does state:

> Tagging [i.e. clicking on] frames in the finest level 76 results in playback of the video; for instance if the j-th keyframe is tagged at the finest level, frames between the j th and (j+1) st keyframes are played back.

(Ratakonda, col. 5, lines 51-54; see also col. 5, lines 56-59); however, it is noted that there is no teaching or suggestion that "frames between the j th and (j+1) st keyframes" are a shot or segment. On the other hand, Ratakonda teaches to the contrary:

> A major limitation of the above schemes is that they treat all shots equally. In most situations it might not be sufficient to represent the entire shot by just one frame. This leads to the idea of allocating a few keyframes per each shot depending upon the amount of 'interesting action' in the shot.

Ratakonda, col. 1, line 64 to col. 2, line 1

> Given total number of keyframes (user specified) 40, <u>each shot is assigned a number of keyframes 42</u>

> depending upon the "action" within the shot, according to well known techniques.

(Ratakonda, col. 4, lines 54-57; emphasis added; also see Ratakonda Figure 2). Ratakonda also states:

> The number of keyframes allocated to a particular shot 's', block 42, is proportionate to the relative amount of cumulative action measure within that shot.

(Ratakonda, col. 6, lines 42-44; see also pruning of keyframes of some shots, Ratakonda, col. 8, lines 31-34)

Figure 5 is in accord with these quotes. Figure 5 shows a coarse level 74 of keyframes, each of which is associated with a group of finest level keyframes. One of the groups has three finest level keyframes. The other has two, that is, a pair. Figure 5 thus shows two groups of different sizes, not pairs. (See also Ratakonda, col. 5, lines 49-51.) Ratakonda does discuss use of "pairwise" clustering algorithms, but this clustering is of finest level keyframes not segments (shots). (Ratakonda, col. 9, lines 40-64) There is, thus, no teaching or suggestion of extracting a feature set of features of each pair of segments.

For at least these reasons, Appellants respectfully submit that Claim 1 is patentable over the cited references, taken separately or in any proper combination.

The other independent Claims 7, 10, 11, 17, 18, 20, 21, 27, and 29 include the same or similar features to those described above in connection with Claim 1 and are believed to be patentable for at least the same reasons. The claims dependent upon these independent claims also are submitted to be patentable for at least the same reasons.

Regarding All Claims: The cited references do not teach or suggest merging video segments with a merging criterion that applies a probabilistic analysis of inter-segment features of pairs of segments.

Claim 1 requires:

> d) merging video segments with a merging criterion that applies a probabilistic analysis to the features of

> the feature set, thereby generating a merging sequence representing the video structure.

The rejection proposes that preparation of textual summaries in Qian 454 meets this language:

> merging the video segments by applying a probabilistic analysis to the extracted set to represent the video structure ("each shot is summarized 16 ...events 22 are inferred from the shot summaries by a domain specific event inference model". Column 3, lines 6-8).

Qian 454 contradicts the rejection. Qian 454, in discussing an animal hunt example, describes various items of "hunt information" and states:

> This "hunt information" is designated true (1) or false (0) and is used in the event inference module to determine whether a valid hunt has been detected.'

(Qian 454, col. 11, lines 48-50; also generally see col. 11, line 7 to col. 12, line 9). This is not a probabilistic analysis. Ratakonda does not add to the teachings of Qian 454 in relation to this feature.

Ratakonda does not teach merging segments, but rather clustering <u>keyframes</u>. (Ratakonda, col. 9, lines 40-64) As a part of the process, Ratakonda also prunes keyframes of some shots. (Ratakonda, col. 8, lines 31-34) As noted above, a segment can have multiple keyframes. Ratakonda indicates:

> The number of keyframes allocated to a particular shot 's', block 42, is proportionate to the relative amount of cumulative action measure within that shot.'

(Ratakonda, col. 6, lines 42-44). This is unlike Claim 1 and contrary to the rejection.

For at least these reasons, Appellants respectfully submit that Claim 1 is patentable over the cited references, taken separately or in any proper combination.

The other independent Claims 7, 10, 11, 17, 18, 20, 21, 27, and 29 include the same or similar features to those described above in connection with Claim 1 and are believed to be patentable for at least the same reasons. The

claims dependent upon these independent claims also are submitted to be patentable for at least the same reasons.

Regarding Claims 1, 21, and 27: The cited references do not teach or suggest metrics of temporal separation between segments and accumulated duration of the segments.

Claim 1 requires "said inter-segment temporal relationship feature including metrics of temporal separation between the segments of the respective said pair and accumulated duration of the segments of the respective said pair". The rejection proposes that Qian 454 teaches extracting a feature set by processing pairs of segments:

> for their visual dissimilarity and temporal relationship, generating a feature including metrics of temporal separation between segments of the respective pair and accumulated duration of segments of the pair (temporal and spatial phenomena), and merging the video segments by applying a probabilistic analysis to the extracted set to represent the video structure ("each shot is summarized 16 ...events 22 are inferred from the shot summaries by a domain specific event inference model". Column 3, lines 6-8).

Qian 454 teaches shot summaries in the form of text descriptors. Some of the descriptors are labeled ""temporal descriptors". Qian 454 states: 'Temporal descriptors represent motion information related to objects and the temporal relations between them. These may be expressed in temporal prepositions, such as, "while", "before", "after", etc.' (Qian 454, col. 11, lines 14-18; emphasis added).

Spatial descriptors similarly represent location and size information related to objects. (Qian 454, col. 11, lines 11-14) The objects are named by object descriptors: 'The object descriptors indicate the existence of certain objects in the video frame; for example, "animal", "tree", "sky/cloud", "grass", "rock", etc.' (Qian 454, col. 11, lines 8-11)

The descriptors of Qian 454 summarize the content of a shot. The metrics of Claim 1 relate to segments not objects: "metrics of temporal separation between the segments of the respective said pair and accumulated duration of the segments of the respective said pair".

The disclosed temporal descriptors: "while", "before", and "after" are not metrics indicating a temporal separation or accumulated duration.

Qian 454 also discloses "Hunt Information" in Figure 9, but teaches that this binary (true/false) information is only used to determine whether a valid hunt has been detected. (Qian 454, col. 11, lines 42-50) Ratakonda does not add to the teachings of Qian 454 in regard to the metrics.

For at least these reasons, Appellants respectfully submit that Claim 1 is patentable over the cited references, taken separately or in any proper combination.

Independent Claims 21 and 27 include the same or similar features to those described above in connection with Claim 1 and are believed to be patentable for at least the same reasons. The claims dependent upon these independent claims also are submitted to be patentable for at least the same reasons.

The rejection fails to state a prima facie rejection, since the rejection fails to address rebuttal evidence showing motivation to not combine the references.

The rejection argues as motivation for one or ordinary skill in the art to combine Qian 454 and Ratakonda:

> It would have been obvious to one of ordinary skill in the art, having the teachings of Qian et al. and Ratakonda before him at the time the invention was made, to modify the segment generation and merging techniques taught by Qian et al. to include the processing of each pair of segments of Ratakonda, in order to obtain not only frames, but also inter-segment similarity processing. One would have been motivated to make such a combination because layered hierarchical structure would have been obtained, as taught by Ratakonda.

(The rejection appears to confuse "segments" and "frames" in arguing motivation for the rejection. Contrary to this statement, frames are not obtained from segment gencrating and merging techniques. A shot or segment is comprised of a plurality of individual frames. (See, e.g., Qian 454, col. 3, lines 37-40) For the sake of clarity, it is assumed that the words "frames" in the preceding quote from the rejection should have been "segments".)

In Appellants' response filed on or about Mar. 6, 2006, Appellants presented evidence in the cited references, denying the proposed motivation and instead providing motivation for onc of skill in the art to not combinc those references. The Office Action responded: "In response to the arguments regarding motivation to combine Qian and Ratakonda the examiner disagrees. Ratakonda clearly teaches layers of a parent/child hierarchy structure and describes the advantages of such structure (Col. 1, line 64 et seq.)."

This citation does not answer the evidence in the cited references that teaches against the cited combination of references. The citcd portion of Ratakonda describes the use of multiple keyframcs to summarize a shot:

> A major limitation of the above schemes is that they treat all shots equally. In most situations it might not be sufficient to rcpresent the entire shot by just one frame. This leads to the idea of allocating a few keyframes per each shot depending upon the amount of "interesting action" in thc shot. The current statc of the art video browsing systems thus split a video sequence into its component shots and represent each shot by a few representative keyframes, where the representation is referred to as "the summary". (Ratakonda, col. 1, line 64 to col. 2, line 1)

How would this be an advantage to a combination of Qian 454 with Ratakonda, unless the combination used kcyframes? Such a combination would not teach or suggest the extracting and merging steps of Claim 1.

In addition, a combination of Qian 454 with Ratakonda is taught against by those references. Ratakonda teaches a "summary" in the form of keyframes of the video. (Ratakonda, col. 2, lines 13-17) In the hierarchy of Ratakonda, each level has less of the keyframes:

> "Video summarization" refers to determining the most salient frames of a given video sequence that may be used as a representative of the video. A method of hierarchical summarization is disclosed for constructing a hierarchical summary with multiple levels, where levels vary in terms of detail (i.e., number of frames). The coarsest, or most compact, level provides the most salient frames and contains the least number of frames. (Ratakonda, col. 2, lines 27-33)

This "summary" of Ratakonda is not compatible with the "summary" of Qian 454, which is textual and lacks details:

> Each shot detected or forced at the first level 4 of the video content analysis technique is summarized 16 at the second level 12 of the technique. <u>The shot summaries provide a means of encapsulating the details</u> of the feature and motion analysis performed at the first 4 and second 12 levels of the technique so that an event inference module in the third level 18 of the technique may be developed independent of the details in the first two levels. <u>The shot summaries also abstract the lower level analysis results so that they can be read and interpreted more easily by humans. This facilitates</u> video indexing, retrieval, and browsing in video databases and the development of algorithms to perform these activities. (Qian 454, col. 10, line 63 to col. 11, line 6; <u>emphasis added</u>)

Ratakonda teaches a hierarchy, in which every level contains detail in the form of one or more keyframes. Qian 454 teaches to the contrary that detail is to be encapsulated in text, which can be read and interpreted more easily by humans.

<u>*Claims 3-5 and 23-26 are patentable as depending from Claim 1 and as follows.*</u>

> As to Claim 3, the rejection stated:
>
> As in Claim 3, Qian et al. teaches obtaining unstructured video frames, generating segments from the shot boundaries based on the color dissimilarity between consecutive frames, extracting a set by processing pairs of segments for their visual dissimilarity and temporal relationship by

> generating color histograms from the consecutive frames and from the histograms, generating a difference signal, thresholding of this signal based on a mean dissimilarity over several frames to produce a signal representative of the existence of a shot boundary (See Claim 23 rejection supra) and merging the video segments by applying a probabilistic analysis to the extracted set to represent the video structure (See Claim 1 rejection supra) and the difference signal to be based on a mean dissimilarity over several frames centered on one frame.

Claim 3 requires a difference signal is based on a mean dissimilarity determined <u>over a plurality of frames centered on one of the consecutive frames</u>. The rejection proposes that Qian 454 teaches: "thresholding of this signal based on a mean dissimilarity <u>over several frames</u>" (<u>emphasis added</u>)

Appellants respectfully traverse. Claim 3 requires a dissimilarity determined over <u>a plurality of frames centered on one of the consecutive frames</u>. Qian 454 teaches use of a difference in histograms between <u>a pair</u> of frames. Qian 454 states:

> A color histogram technique may be used to detect the boundaries of the shots. The difference between the histograms <u>of two frames</u> indicates a difference in the content of those frames. When the difference between the histograms <u>for successive frames</u> exceeds a predefined threshold, the content of the <u>two frames</u> is assumed to be sufficiently different that the frames are from different video shots. Other known techniques could be used to detect the shot boundaries. (Qian 454, col. 3, lines 40-50; <u>emphasis added</u>)

The combination of the cited references adds nothing to Qian 454, since Ratakonda is similar to Qian 454:

> Image color histograms, i.e., color distributions, constitute representative feature vectors of the video frames and are used in shot boundary detection 38 and keyframe selection. Shot boundary detection 38 is performed using a threshold method, where differences between histograms of <u>successive frames</u> are compared. (Ratakonda, col. 4, lines 48-54; <u>emphasis added</u>)

The emphasized language in the above quotes contradicts the rejection.

The Office Action stated in relation to Claim 5: "As in Claim 5, Qian et al. teaches computing a mean color histogram for each segment and a visual dissimilarity feature metric from the difference between mean color histograms for pairs of segments (Column 3, lines 42-50 and Figure 5)." "In response to the arguments regarding claim 5 and 6, Ratakonda teaches processing each pair of segments for dissimilarity in the same way Qian does for frames as seen supra."

Claim 5 requires computing a mean color histogram for each segment and computing a visual dissimilarity feature metric from the difference between mean color histograms for pairs of segments. Since Claim 5 depends from Claim 1, it also requires generating video segments from the unstructured video by detecting shot boundaries based on color dissimilarity between consecutive frames. There are, thus, two different types of dissimilarity features required by Claim 5: a first type is between consecutive frames and a second type is between pairs of segments. There is only one such feature in Qian 454. This is apparent from the office action, which cites the same portion of Qian 454 for both Claim 1 (Column 3, lines 42-43) and Claim 5 (Column 3, lines 42-50 and Figure 5). (Figure 5 of Qian 454 relates to a "sample mean" that is unrelated to the subject matter of Claim 5. See Qian 454, col. 6, lines 16-18)

Appellants respectfully
take issue with the examiner's statement in the rejection that:

> In response to the arguments regarding claim 5 and 6, Ratakonda teaches processing each pair of segments for dissimilarity in the same way Qian does for frames as seen supra.

Ratakonda and Qian 454 both teach a color dissimilarity feature <u>for a pair of frames not a pair of segments</u>. Ratakonda states:

> Image color histograms, i.e., color distributions, constitute representative feature vectors of the video frames and are used in shot boundary detection 38 and keyframe selection. Shot boundary detection 38

> is performed using a threshold method, where differences between histograms of <u>successive frames</u> are compared. (Ratakonda, col. 4, lines 48-54)

The cited combination of references only teaches determining a color dissimilarity feature between <u>frames</u>.

Claim 24 has language like Claim 28 and is also patentable on the same grounds as that claim (discussed below). Claims 25-26 have language taken from Claims 7-8 (discussed below) and are also patentable on the same grounds as those claims.

<u>Independent Claim 7 is Patentable in its Own Right</u>

> The rejection stated in relation to Claim 7.
>
> As in Claim 7, Qian et al. teaches the method of claim 1 as seen supra, and generating parametric mixture models (summaries created by shot summarization 16, Figure 1) to represent class-conditional densities of inter-segment features (based on temporal information and color analysis, See Claim 1 rejection supra) of the feature set, parametric mixture models being statistical models (Col. 3, lines 34-35, and Col. 4, lines 30 et seq.) and applying the merging criterion to the parametric mixture models (event inference 20/detected events 22, Figure 1).

Claim 7 requires "extracting a feature set by processing pairs of said segments". As discussed in relation to Claim 1, the cited references not only do not teach or suggest, but rather teach against extracting inter-segment features by processing pairs of segments. In Qian 454, shots are compared in the form of summaries. Each of the individual shots, in Qian 454, are summarized with descriptors, such as "animal" and "tree", and the descriptors of different shots are compared, but not in pairs. (Qian 454, col. 10, line 61 to col. 12, line 9) Qian 454 teaches against comparisons between shots based upon "details" and teaches against presentation of image content to users. Qian 454 instead presents summaries to be read and interpreted. (Qian 454, col. 10, line 63 to col. 11, line 6) As discussed above,

Figures 1 and 5 and related text of Ratakonda does not teach or suggest "extracting a feature set by processing pairs of said segments". Figure 5 of Ratakonda does not represent a comparison of shots, with each shot having a representative keyframe, since Ratakonda teaches against use of one keyframe per shot:

> "In most situations it might not be sufficient to represent the entire shot by just one frame." (Ratakonda, col. 1, lines 65-66)

> 'Given total number of keyframes (user specified) 40, each shot is assigned a number of keyframes 42 depending upon the "action" within the shot, according to well known techniques.' (Ratakonda, col. 4, lines 54-57; emphasis added; also see Ratakonda Figure 2)

The cited references teach nor more together then they do individually.

The rejection also proposes that "said parametric mixture models being statistical models" is taught at Qian 454, col. 3, lines 34-35, and col. 4, lines 30 et seq. The first citation is presumed to be a phrase in Qian 454, col. 3, at lines 33-35, which states: "but the technique may be easily extended to other domains by adding modules containing rules specific to the additional domain."

On its face, this language appears to present an inappropriate rejection in that it conveys the meaning that the claimed invention would have been well within the ordinary skill of the art at the time the claimed invention was made, because the references relied upon teach that all aspects of the claimed invention were individually known in the art. (Such art is not sufficient to establish a prima facie case of obviousness without some objective reason to combine the teachings of the references. MPEP 2143.01) The second citation from Qian 454 (col. 4, lines 30 et seq.) is indefinite in length, but is believed to refer to Qian 454, col. 4, lines 32-60, which discuss use of a five level pyramid technique to estimate global motion. (See Qian 454, col. 4, lines 18-34)

The combination of the two citations, Qian 454, col. 3, lines 34-35, and col. 4, lines 30 et seq., is understood to represent the position that a "domain"

of Qian 454, col. 3, lines 34-35 could be the five level pyramid technique used to estimate global motion at Qian 454, col. 4, lines 30 et seq. There is no support for this argument in the cited references.

Qian 454 uses the term "domain" in a specific sense that is contrary to the rejection. Qian 454 states:

> The third category of techniques for analyzing video content applies rules relating the content to features of a specific video domain or content subject area. For example, methods have been proposed to detect events in football games, soccer games, baseball games and basketball games. The events detected by these methods are likely to be semantically relevant to users, but these methods are heavily dependent on the specific artifacts related to the particular domain, such as editing patterns in broadcast programs. This makes it difficult to extend these methods to more general analysis of video from a broad variety of domains.
>
> What is desired, therefore, is a method of video content analysis which is adaptable to reliably detect semantically significant events in video from a wide range of content domains. (Qian 454, col. 1, lines 48-62; emphasis added)
>
> As a result, the technique of the present invention detects event which are meaningful to a video user and the technique may be extended to a broad spectrum of video domains by incorporating shot summarization and event inference modules that are, relatively, specific to the domain or subject area of the video which operate on data generated by visual analysis processes which are not domain specific. (Qian 454, col. 2, lines 7-18; emphasis added)
>
> Event inference 20 is based on domain or subject matter specific knowledge developed from observation of video and shot summaries generated at the intermediate level 12 of the technique. For example, an animal hunt usually comprises an extended period during which the animal is moving

> fast, followed by the slowing or stopping of the animal. (Qian 454, col. 11, lines 52-58; emphasis added)

> At the third level of the process, event inference modules provide the domain specific structure necessary for reliable event detection, but the technique may be easily extended to other domains by adding modules containing rules specific to the additional domain. (Qian 454, col. 3, lines 31-35; emphasis added)

(Qian 454 at col. 9, lines 15-22, also mentions the spatial domain relative to the spatial-frequency decompositions of Gabor filters. The rejection's proposed use of "domain" does not meet this alternative definition. In the following discussion, the " spatial domain" is not considered.)

As the emphasized language in the above quotes indicates, the term "domain" in Qian 454 is the subject area of the content of a video. An example of a domain, discussed extensively in Qian 454, is an animal hunt. The five level pyramid technique used to estimate global motion in Qian 454 is not an example of a domain.

Qian 454 also specifically teaches that the five level pyramid technique is not specific to a particular domain. Qian 454 states:

> At the lowest level of the technique, visual analysis processes which are not specific to an application or video domain provide basic information about the content of the video. (Qian 454, col. 3, lines 17-20; emphasis added; see also col. 2, lines 11-18)

> At the first level 4 of the technique, the global motion of the video content is estimated 8 for each pair of frames in a shot. (Qian 454, col. 3, lines 59-61; emphasis added)

> In the instant technique, global motion is estimated with a five level pyramid technique. (Qian 454, col. 4, lines 18-19; emphasis added)

The emphasized language shows, the position taken in the rejection is unsupported by the cited references, since the five level pyramid technique for estimating global motion is not domain specific.

The rejection proposes that creating summaries in Qian 454 by shot summarization based on temporal information and color analysis, corresponds to generating parametric mixture models to represent class-conditional densities of inter-segment features of a feature set extracted by processing pairs of segments. The rejection also proposes that applying the merging criterion to the parametric mixture models is taught in Qian 454 by "(event inference 20/detected events 22, Figure 1)". The cited combination of references do not disclose these features.

The cited references do not teach or suggest generating parametric mixture models to represent class-conditional densities of inter-segment features of the feature set. Claim 7 requires generating "parametric mixture models" that are defined by the specification and usage in the art as types of statistical models. (See application page 4, lines 25-30; page 13, lines 14-29; also see U.S. Patent No. 5,710,833.) As discussed above, the cited references do not teach or suggest this. The rejection's "summaries created by shot summarization" are not statistical models. Qian 454 teaches summaries, in which shot descriptors are described as indicating as to a particular shot: "the existence of certain objects", "location and size information related to objects and the spatial relations between objects", and "motion information related to objects and the temporal relations between them". (Qian 454, col. 11, lines 9, 11-13, and 15-16; see also the above discussion of summarization.)

In Claim 7, the parametric mixture models are generated to represent class-conditional densities of inter-segment features. The cited references, as noted above, fail to teach or suggest the processing of pairs of segments to provide the features of the feature set. Shot descriptors for each segment are taught by Qian 454. (See Qian 454, col. 10, lines 61-62: "Each shot ... is summarized"). Ratakonda teaches a "summary" in the form of a hierarchy of

keyframes, which is not limited to one keyframe per shot. (Ratakonda, Figures 1 and 5, col. 3, lines 30-45; col.5, lines 44-63; and col. 13, lines 22-31; col. 1, lines 65-66; col. 4, lines 54-57)

The rejection is not supported by the cited references and must be withdrawn.

Claims 6 and 8 are allowable as depending from Claim 7 and as follows.

Claim 6 is allowable on the same grounds as discussed above in relation to similar language in Claim 1. Claim 6 requires that the processing of pairs of segments includes processing for a temporal separation between pairs of segments and for an accumulated temporal duration of pairs of segments. The language relied upon in the rejection (Qian 454, col. 3, lines 6-8) relates to inferences based on textual shot summaries. Qian et al. states: "The shot summaries also abstract the lower level analysis results so that they can be read and interpreted more easily by humans." (Qian 454, col. 11, lines 1-3).

Qian 454 teaches summarization that encapsulates the details of the feature and motion analysis of each shot using descriptors. (Qian 454, col. 10, line 63 to col. 11, line 8) The domain specific event inference model uses the descriptors. (Qian 454, col. 11, lines 51-55) Events are inferred by matching the occurrence of objects and their spatial and temporal relationships detected in each of the shots. (Qian 454, col. 12, lines 6-7; generally see col. 11, line 58 to col. 12, line 9) Examples of shot descriptors are provided:

> In general, shot descriptors used in the shot summary include object, spatial, and temporal descriptors. The object descriptors indicate the existence of certain objects in the video frame; for example, "animal", "tree", "sky/cloud", "grass", "rock", etc. The spatial descriptors represent location and size information related to objects and the spatial relations between objects in terms of spatial prepositions, such as "inside", "next to", "on top of", etc. Temporal descriptors represent motion information related to objects and the temporal relations between them. These may be expressed in temporal prepositions, such as, "while", "before", "after," etc. (Qian 454, col. 11, lines 7-18)

Qian 454 does not teach descriptors of temporal separation between pairs of segments and/or for accumulated temporal duration between pairs of segments. In Qian 454, temporal descriptors represent motion information related to objects in a segment and the temporal relations between those objects in that segment. This is unlike Claim 6, which requires processing pairs of segments for a temporal separation between pairs of segments and for an accumulated temporal duration of pairs of segments.

Ratakonda does not disclose the features of Claim 6. Ratakonda instead teaches the use of histogram clustering to determine keyframes. (See Ratakonda, col. 9, lines 30-53, quoted above)

Regarding Claim 8, the Office Action stated in relation to Claim 8: 'As in Claim 8, it is notoriously well known that queues are used to implement hierarchical displays. The examiner takes official notice of this teaching. It would be obvious to one of ordinary skill in the art to combine the use of the organizing video segments into hierarchies with a queue implementation.'

The Office Action presents a different rejection of Claim 15, which is very similar to Claim 8: "As in Claim 15, US Patent 6721454 and Ratakonda teach performing the merging in a hierarchical queue by initializing the queue by introducing each feature in the queue with a priority of the probability of merging each corresponding pair of segments, depleting the queue by merging the segments if the criterion is met, and updating the queue based on the updated model (See Claim 8 rejection supra)."

The Office Action also stated: 'In response to the arguments regarding claim 8 and 15, Qian teaches the process of "inserting" merges frames together, constituting a pair of segments that define the event and updating the model of the merged segment. Ratakonda further illustrates step d as seen supra. Furthermore, claim 15 is interpreted with respect to the official notice of Claim 8. Claim 15 merging and depleting sequence can further be illustrated by Qian figures 1 and 7 with corresponding text.'

In the amendment filed on or about March 6, 2006, in relation to Claim 8, Appellants presented a demand for clarification of the official notice stating:

> "Clarification of the rejections of Claims 8 and 15 is requested, particularly as to the metes and bounds of the official notice taken and of the relied upon teachings of Qian 454 and Ratakonda."

The Office Action presents a response that is inadequate. The metes and bounds of the official notice taken are not addressed. Appellants have insufficient information to prepare a response to the rejection. The rejection will not stand.

The Office Action does present additional argument as to the references, but in so doing again confuses the meanings of "frames" and "segments", stating:

> "Qian teaches the process of "inserting" merges <u>frames</u> together, constituting a pair of <u>segments</u> that define the event and updating the model of the merged <u>segment</u>." (<u>emphasis added</u>)

As discussed above, frames are not segments. Thus, merging frames together cannot constitute "a pair of segments that define the event and updating the model of the merged segment."

The Office Action indicates "Qian teaches the process of "inserting" merges frames together". It is unclear what is meant by this. Qian 454 mentions "inserting", but this term has the same meaning as "forcing", that is, adding another shot boundary into a sequence of frames. Qian 454 states:

> In addition to the shot boundaries detected in the video sequence, shot boundaries may be forced or <u>inserted</u> into the sequence whenever the global motion of the content changes. As a result, the global motion is relatively homogeneous between the boundaries of a shot. In addition, shot boundaries may be forced after a specific number of frames (e.g., every 200 frames) to reduce the likelihood of missing important events within extended shots." (Qian 454, col. 3, lines 51-58; <u>emphasis added</u>)

Inserting a shot boundary divides a video sequence. This is the opposite of merging.

The Office Action also states: "Ratakonda further illustrates step d as seen supra." This is understood to mean the same thing as the discussion of step d in the rejection of Claim 1: "Ratakonda teaches a video event detection and segmentation merging method similar to that of Qian et al."

As discussed above, Ratakonda does not teach merging segments, but rather producing a hierarchy of keyframes by clustering <u>keyframes</u>. (Ratakonda, col. 9, lines 40-64) As a part of the process, Ratakonda also prunes keyframes of some shots. (Ratakonda, col. 8, lines 31-34) A segment can have multiple keyframes. (Ratakonda, col. 6, lines 42-44)

The rejection of Claim 8 is otherwise limited to the words of the official notice. The rejection states that it is notoriously well known that queues are used to implement hierarchical displays. This statement addresses only one phrase of Claim 8: "performed in a hierarchical queue" and does not teach or suggest the steps of:

> initializing the queue by introducing each feature into the queue with a priority equal to the probability of merging each corresponding pair of segments;
>
> depleting the queue by merging the segments if the merging criterion is met; and
>
> updating the model of the merged segment and then updating the queue based upon the updated model.

<u>B. Claims 9 and 11-22 are patentable under 35 U.S.C. 103(a) over Qian et al., US Patent 6721454 (hereafter (Qian 454) and Ratakonda, US Patent 5956026 and further in view of Qian et al., US Patent 6616529 (hereafter Qian 529).</u>

Claim 9 is patentable as depending from Claim 1.

Claim 11 is patentable as discussed above in relation to Claim 7.

Claims 12-16 are patentable as depending from Claim 11. Claims 12-13 are also patentable on the same basis as Claims 5-6, respectively. Claim 15 is patentable on the same basis as Claim 8.

Claims 17-18 are patentable as discussed above in relation to Claim 7.

Claim 19 is patentable as depending from Claim 18.

Claim 20 is patentable as discussed above in relation to Claim 7.

Claim 21 is patentable on the same basis as Claim 1.

Claim 22 is patentable as depending from Claim 21.

IX. Conclusion

For the above reasons, Appellants respectfully request that the Board of Patent Appeals and Interferences reverse the rejection by the Examiner and mandate the allowance of Claims 1 and 3 – 29..

Respectfully submitted,

Attorney for Appellants
Registration No. 52,118

Telephone: (585) 726-7522
Facsimile: (585) 477-4646
Enclosures

If the Examiner is unable to reach the Appellant(s) Attorney at the telephone number provided, the Examiner is requested to communicate with Eastman Kodak Company Patent Operations at (585) 477-4656.

X. Appendix I - Claims on Appeal

1. A method for structuring video by probabilistic merging of video segments, said method comprising the steps of:

a) obtaining a plurality of frames of unstructured video;

b) generating video segments from the unstructured video by detecting shot boundaries based on color dissimilarity between consecutive frames;

c) extracting a feature set by processing pairs of said segments, said extracting generating an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments, said inter-segment temporal relationship feature including metrics of temporal separation between the segments of the respective said pair and accumulated duration of the segments of the respective said pair; and

d) merging video segments with a merging criterion that applies a probabilistic analysis to the features of the feature set, thereby generating a merging sequence representing the video structure.

2 (cancelled).

3. The method as claimed in claim 23 wherein the difference signal is based on a mean dissimilarity determined over a plurality of frames centered on one of the consecutive frames.

4. The method as claimed in claim 23 further including the step of morphologically transforming the threshold difference signal with a pair of structuring elements that eliminate the presence of multiple adjacent shot boundaries.

5. The method as claimed in claim 1 wherein the processing of pairs of segments for visual dissimilarity in step c) comprises the steps of computing a mean color histogram for each segment and computing a visual dissimilarity feature metric from the difference between mean color histograms for pairs of segments.

6. The method as claimed in claim 7 wherein the processing of pairs of segments for their temporal relationship in step c) comprises the processing of pairs of segments for a temporal separation between pairs of segments and for an accumulated temporal duration of pairs of segments.

7. A method for structuring video by probabilistic merging of video segments, said method comprising the steps of:

a) obtaining a plurality of frames of unstructured video;

b) generating video segments from the unstructured video by detecting shot boundaries based on color dissimilarity between consecutive frames:

c) extracting a feature set by processing pairs of said segments, said extracting generating an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments; and

d) merging video segments with a merging criterion that applies a probabilistic analysis to the features of the feature set, thereby generating a merging sequence representing the video structure;

wherein step d) comprises the steps of:

generating parametric mixture models to represent class-conditional densities of inter-segment features of the feature set, said parametric mixture models being statistical models; and

applying the merging criterion to the parametric mixture models.

8. The method as claimed in claim 7 wherein step d) is performed in a hierarchical queue and comprises the steps of:

initializing the queue by introducing each feature into the queue with a priority equal to the probability of merging each corresponding pair of segments;

depleting the queue by merging the segments if the merging criterion is met; and

updating the model of the merged segment and then updating the queue based upon the updated model.

9. The method as claimed in claim 1 wherein representing the merging sequence is represented in a hierarchical tree structure.

10. A computer storage medium having instructions stored therein for causing a computer to perform the acts of:

generating video segments from unstructured video by detecting shot boundaries based on color dissimilarity between consecutive frames;

extracting a feature set by processing pairs of segments, said extracting generating an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments; and

merging video segments with a merging criterion that applies a probabilistic analysis to the fcatures of the feature set, thereby generating a merging sequencc representing the video structure;

wherein said merging further comprises the steps of:

generating statistical models of the feature set; and

applying the merging criterion to the statistical models.

11. A method for structuring video by probabilistic merging of video segments, said method comprising the steps of:

a) obtaining a plurality of frames of unstructured vidco;

b) generating video segments from the unstructured video by detecting shot boundaries based on color dissimilarity between consecutive video frames;

c) extracting a feature set by processing pairs of segments, said extracting generating an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments;

d) generating a parametric mixture model of the inter-segment features comprising the feature set, said parametric mixture model being a statistical model; and

e) merging video segments with a merging criterion that applies a probabilistic Bayesian analysis to the parametric mixture model, thereby generating a merging sequence representing the video structure.

12. The method as claimed in claim 11 wherein the processing of pairs of segments for visual dissimilarity in step c) comprises the steps of computing a mean color histogram for each segment and computing a visual dissimilarity feature metric from the difference between mean color histograms for pairs of segments.

13 . The method as claimed in claim 11 wherein the processing of pairs of segments for their temporal relationship in step c) comprises the processing of pairs of segments for a temporal separation between pairs of segments and for an accumulated temporal duration of pairs of segments.

14 . The method as claimed in claim 11 wherein the parametric mixture model generated in step d) represents class-conditional densities of the inter-segment features comprising the feature set.

15 . The method as claimed in claim 11 wherein step e) is performed in a hierarchical queue and comprises the steps of:

initializing the queue by introducing each feature into the queue with a priority equal to the probability of merging each corresponding pair of segments;

depleting the queue by merging the segments if the merging criterion is met; and

updating the model of the merged segment and then updating the queue based upon the updated model.

16 . The method as claimed in claim 11 wherein the merging sequence is represented in a hierarchical tree structure that includes a frame extracted from each segment and displayed at each node of the tree.

17. A computer storage medium having instructions stored therein for causing a computer to perform acts for structuring video by probabilistic merging of video segments, the acts including:

obtaining a plurality of frames of unstructured video;

generating video segments from the unstructured video by detecting shot boundaries based on color dissimilarity between consecutive video frames;

extracting a feature set by processing pairs of segments, said extracting generating an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments;

generating a parametric mixture model of the inter-segment features comprising the feature set, said parametric mixture model being a statistical model; and

merging video segments with a merging criterion that applies a probabilistic Bayesian analysis to the parametric mixture model, thereby generating a merging sequence representing the video structure.

18. A method for structuring video by probabilistic merging of video segments, said method comprising the steps of:

a) obtaining a plurality of frames of unstructured video;

b) generating video segments from the unstructured video by detecting shot boundaries based on color dissimilarity between consecutive video frames;

c) extracting a feature set by processing pairs of segments, said extracting generating an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments;

d) merging adjacent video segments with a merging criterion that applies a probabilistic Bayesian analysis to parametric mixture models derived from the feature set, said parametric mixture models being statistical models, thereby generating a merging sequence; and

e) representing the merging sequence in a hierarchical tree structure.

19 . The method as claimed in claim 18 wherein representing the merging sequence in a hierarchical tree structure includes displaying a frame extracted from each segment.

20. A computer storage medium having instructions stored therein for causing a computer to perform probabilistic merging of video segments, said instructions performing the acts of:

a) obtaining a plurality of frames of unstructured video;

b) generating video segments from the unstructured video by detecting shot boundaries based on color dissimilarity between consecutive video frames;

c) extracting a feature set by processing pairs of segments, said extracting generating an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments;

d) merging adjacent video segments with a merging criterion that applies a probabilistic Bayesian analysis to parametric mixture models derived

from the feature set, said parametric mixture models being a statistical models, thereby generating a merging sequence; and

e) representing the merging sequence in a hierarchical tree structure.

21. A method for structuring video by probabilistic merging of video segments, said method comprising:

generating video segments from an unstructured plurality of video frames by detecting shot boundaries based on color dissimilarity between consecutive frames;

extracting a feature set by processing pairs of segments, said extracting generating an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments, said inter-segment temporal relationship feature including metrics of temporal separation between the segments of the respective said pair and accumulated duration of the segments of the respective said pair;

merging the video segments with a merging criterion that applies a probabilistic analysis to the feature set, thereby generating a merging sequence representing the video structure, the merging being independent of any empirical parameter determination; and

generating a hierarchy with the merged video segments, the hierarchy having a merging sequence represented by a binary partition tree.

22 . The method as claimed in claim 21 wherein the merging the video segments includes maximizing the a posteriori probability mass function of a binary random variable that represents inter-segment features of the video segments.

23 . The method as claimed in claim 1 wherein step b) comprises the steps of:

generating color histograms from the consecutive frames;

generating a difference signal from the color histograms that represents the color dissimilarity between consecutive frames; and

thresholding the difference signal based on a mean dissimilarity determined over a plurality of frames, thereby producing a signal that indicates an existence of a shot boundary.

24 . The method as claimed in claim 1 wherein said extracting of said inter-segment temporal relationship feature of each said pair of segments including determining a number of frames separating the respective said pair of segments and determining an accumulated number of frames in said segments of the respective said pair of segments.

25 . The method as claimed in claim 1 wherein step d) comprises the steps of:

generating parametric mixture models to represent class-conditional densities of inter-segment features of the feature set, said parametric mixture models being statistical models; and

applying the merging criterion to the parametric mixture models.

26 . The method as claimed in claim 25 wherein step d) is performed in a hierarchical queue and comprises the steps of:

initializing the queue by introducing each feature into the queue with a priority equal to the probability of merging each corresponding pair of segments;

depleting the queue by merging the segments if the merging criterion is met; and

updating the model of the merged segment and then updating the queue based upon the updated model.

27. A method for structuring video by probabilistic merging of video segments, said method comprising the steps of:

generating video segments from a plurality of frames of unstructured video by detecting shot boundaries based on color dissimilarity between consecutive frames;

computing an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments, said inter-

segment temporal relationship feature including metrics of temporal separation between the segments of the respective said pair and accumulated duration of the segments of the respective said pair; and

d) merging video segments with a merging criterion that applies a probabilistic analysis to said features, thereby generating a merging sequence representing the video structure.

28 . The method of claim 27 wherein said computing of said inter-segment temporal relationship feature of each said pair of segments further comprises determining a number of frames separating the respective said pair of segments and determining an accumulated number of frames in said segments of the respective said pair of segments.

29. A method for structuring video by probabilistic merging of video segments, said method comprising the steps of:

obtaining a plurality of frames of unstructured video;

generating video segments from the unstructured video by detecting shot boundaries based on color dissimilarity between consecutive frames;

extracting an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments, said extracting of said inter-segment temporal relationship feature of each said pair of segments including determining a number of frames separating the respective said

pair of segments and determining an accumulated number of frames in said segments of the respective said pair of segments; and

merging video segments with a merging criterion that applies a probabilistic analysis to the features of the feature set, thereby generating a merging sequence representing the video structure.

XI. Appendix II - Evidence

None.

XII. Appendix III – Related Proceedings

None.